2001 年 ベイジアンネットチュートリアル
Tutorial on Bayesian Networks (BN2001)
Tokyo, Japan, July 29, 2001.

不確実性モデリングのための情報表現：ベイジアンネット

Bayesian Netoworks as a representation for uncertainty modeling

本村 陽一*
産業技術総合研究所 情報処理研究部門

Abstract: ベイジアンネットはこれまで様々な分野において，様々な視点から研究が進められてきた．ここでは主に不確実性を扱うための情報表現形式としての観点から解説を行う．とくにベイジアンネットが必要とされる背景，どのようにして利用するのか，適用できる問題領域はどのようなものか，についてそれぞれ述べる．またベイジアンネットは条件付き確率に基づいて不確実性を含む問題領域を記述する枠組を提供している．この点からベイジアンネットの自然な拡張としての連続，非線形領域などへの拡張についても触れる．

1 はじめに

社会全体の情報化，IT 化がますます進み，情報システムが対象とする問題領域も拡大しつつある．20 世紀後半はファジィやニューラルネットなど様々な情報処理パラダイムが提案され，それらについての議論に注目が集まることが多かったようであるが，今後の問題領域の拡大に対してはむしろ情報処理システムが対象とする人間，環境といった複雑な対象をどのように扱えばよいのか，という情報処理を実際に可能にするためのモデリング手法の重要性が増してきているようにも思える．不確実性を含むために事前に確定することのできない問題対象にたいしては，これをうまくモデル化することによって，実行時にありえる可能性を的確に予測し，柔軟に対応することのできる情報処理が期待される．

例えば最近ペットロボットやヒューマノイドロボットなどが脚光をあびているが，このような対話システムは従来のプログラミング技術だけでは適切に動作させることは難しい．対話的なロボットは実環境の中で利用者と接して，状況や文脈に応じて各利用者の望む通りに動作することが必要であるが，この時，設計者が予想もしていなかったような状況で，ノイズに埋もれた音声など不完全な観測情報からでもリアルタイムで適切な動作を行わなければならない．この時，どのようなアクションが適切であるかをシステムが決定するだけの十分な情報を全て揃えることは容易ではないため，完全な情報を仮定し，単に組み込まれたプログラムを順に実行していくだけでは明らかに不十分である[1]．そこで不完全情報のもとでも頑健に動作する知的な制御システムが必要とされ，そのために外部の環境や利用者についての不確実性を反映した内部モデルと予測メカニズムが必要になる．また，それぞれが異なった感性や価値観を持つユーザの意図を汲んで，情報提供を行う検索システムなどへの期待，例えば WWW における顧客管理 (eCRM)，リコメンデーションシステムなどへのニーズも高まってきている．こうした個人の嗜好性や価値観などの要素もやはり非決定的なものであり，不確実性モデリングと予測技術はこうした問題領域でもやはり重要である．さらにデータの発生メカニズム自体は明らかではなくとも，膨大な観測データと傍証的事実が蓄積されてきたことで，そこから重要なモデルを発見することが期待される遺伝子情報処理も今後非常に重要な問題領域である．ここで開発される有効な手法はそれ自体の重要性もさることながら，同様に大規模なデータベースが徐々に揃いつつある様々な分野でデータの発生過程に不確実性を含むモデル化，予測技術，知識構造発見にも期待できる．こうした種類の要請に応えることが今後の情報処理技術に課せられた大きな役割であろう．

一方，古くから不確実性を含む情報は統計学，確率論の枠組によって扱われてきたが，歴史的には小量のデータと比較的単純なモデルによる数理的手法の研究を中心として進められてきた．しかしここでもデータ量と計算パワーの増大によって，より複雑なモデルの開発と実際

*〒 305-8568 茨城県つくば市梅園 1-1-1 つくば中央第二，tel: 0298-61-5836, e-mail: y.motomura@aist.go.jp, URL: http://staff.aist.go.jp/y.motomura/

[1]音声認識誤りの影響から思わぬ動作を実行しオーナーをいらいらさせる AIBO をよく見かける．それでもペットであればそれでもすねた所が可愛いなどと思わせる所がミソなのだが，これがより大型のヒューマノイドロボットであると，可愛いだけではすまされないかも知れない

的な応用へと関心が移りつつあるようである．モデルをより複雑化するための一つの有効な手段は，確率変数の間の相関や条件付き依存関係にグラフ構造を導入するグラフィカルモデリング [39, 44] である．ベイジアンネットもこのグラフィカルモデルの一種で，とくに確率変数間の依存関係を条件付き確率と有向リンクを使ったモデルである．グラフィカルモデル自体は統計分野におけるグラフ構造を持つモデル一般の名称で，ベイジアンネットという名称は主にアメリカを中心とする人工知能分野で主に用いられてきた．他にも Belief network と呼ばれたり，これに特殊なノードを加えた Decision network の他，異なる研究領域で独立して研究されてきた Hidden Markov Model やボルツマンマシンなども広い意味ではベイジアンネットの特殊形として理解することができる．ベイジアンネットを通じて問題をモデル化することで，不確実性に関する性質はこれらの他の確率モデルにも共通して適用でき，具体的な計算アルゴリズムを実行する時に個々のモデル特有の効率化ができるのである．そこで本稿では不確実性モデリングのための情報表現，という視点を中心にしてベイジアンネットを解説する．

2　不確実性モデリング

まずはじめに，完全に観測できない事象を扱う確率的な枠組について見てみよう．例えば，将来の天気や雑音混じりの信号，ユーザの意図のように確定値を得ることが難しいものがある．これらを体系的に取り扱うために確率変数を用いることができる．複雑な要因やノイズの影響などによって不確定さを含む対象を確率変数として大文字 X で表し，その変数がとりえる具体値は小文字 $x_1, x_2, \cdots, x_n$ のように表す．ここで変数 X の確率分布を考えることができ，そのおおまかな傾向はその統計量 (例えば平均や分散, エントロピーなど) によって特徴づけることができる．次に変数間の依存関係を考える．例えば変数 X_i と X_j があり X_i が a という値を取る時には X_j は b となるという関係が成立しているとき，X_j が X_i に依存しているとする (if $X_i = a$ then $X_j = b$)．現実に起きている様々な事象を考えるとこうした変数間の依存関係は複雑であり，「if $X_1 = a_1, \cdots, X_i = a_i, \cdots$, then $X_j = b$」のように明示的に全ての関係を列挙することはあまり現実的でなく，またたとえこのような IF-THEN ルールを膨大に挙げたとしても実際には例外などがあり，必ずしも完全に状況を記述することは難しいだろう．ここでルールが成立するかどうかに関して不確実性を考え，また依存関係の程度を定量的に表すため，「$X_i = a_i$ であるとき $X_j = b$ である確率は $P(X_j = b \mid X_i = a_i)$」という確率的な表現を導入してみよう (これは命題論理表現を確率化したものになっている)．二つの量 x, y の間の一意的な依存関係は，例えば関数 $y = f(x)$ によって表せる．これと同様に，確率変数 X_i,X_j の依存関係は条件付き確率分布 $P(X_j|X_i)$ によって表すことができる．これは X_i のとる値に応じて，X_j の分布が影響をうけ，その依存関係の定量的関係が条件付き確率分布 $P(X_j|X_i)$ で定められることを示している．このような確率的枠組によって不確実な対象をモデル化する不確実性モデリングを行うことで，対象となる問題領域を理解し，計算システムが取り扱えるように記述することができる．このモデルを使って，ある変数がどのような値をとるかについて予測を行ったり，ありえる全ての可能性とその確率を考慮した最適な意志決定が行えるようになる．

3　ベイジアンネット

上で述べた確率的な枠組によって，ある変数の間の不確実のために漠然とした因果関係は定量的な条件付き依存性として表現することができた．さらに因果関係が複雑な問題領域を表すために，複数の変数の間の因果関係を順次結んでいき，グラフ構造を持つネットワークとして表すことを考える．変数間の定性的な依存関係はグラフ構造によって表され，さらに変数間の定量的な依存関係を条件付き確率で表したモデルがベイジアンネットワークである．この変数間の依存関係というものは，本来因果律，状態遷移など様々な相互作用に起因するものであるが，それらを一括して非決定的な条件付き確率に帰着させた所が，ベイジアンネットによるモデル化のポイントである．それによりかなり多くの非決定的な問題領域を記述することができている．一方，その結果を解釈する時には依存関係の意味についてはよく注意[2]する必要がある．

さて，X_i, X_j の間の条件付き依存性をベイジアンネットワークでは向きのついたリンクによって $X_i \to X_j$ と表し，X_i を親ノード，X_j は子ノードと呼ぶ．親ノードが複数あるとき子ノード X_j の親ノードの集合を $\pi(X_j) = \{X_1, \cdots, X_i\}$ と書くことにする．この場合の変数 X_j に関する依存関係は条件付き確率,

$$P(X_j \mid \pi(X_j)) \tag{1}$$

で定義され，これは X_j を子ノード, $\pi(X_j)$ を親ノード群とする木構造になる．さらに n 個の確率変数 $X_1 \cdots, X_n$ があるとき，全ての確率変数の同時確率分布は式 2 のようになり，各子ノードとその親ノード群からなる局所木

[2]事象の発生順序や包含関係など

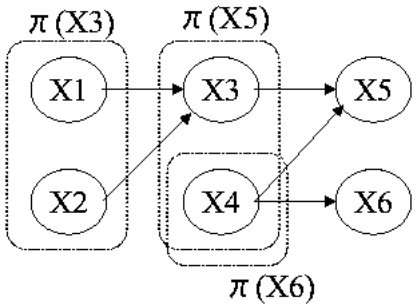

図 1: ベイジアンネットワーク

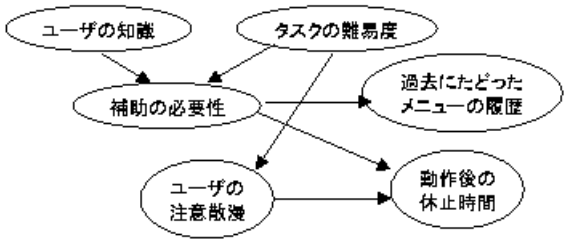

図 2: ベイジアンネットワーク: ユーザモデルの例

を組み合わせたグラフ構造で表せる (1)．

$$P(X_1, \cdots, X_n) = \prod_j P(X_j \mid \pi(X_j)). \qquad (2)$$

つまり，式 2 の左辺の同時確率分布は局所的な木構造に分割した右辺の各項の積として計算される．

この局所木 $\pi(X)$ の集合，すなわちグラフ構造と，各ノードに割り当てた条件付き確率 $P(X_j \mid \pi(X_j))$ の集合によって，ベイジアンネットワークが定義できる．離散変数の場合，子ノードの親ノードに関する条件付き確率は全ての状態における条件付き確率を並べた表，CPT(Conditional Probability Table) によって表す．例えば親ノードがある状態 $\pi(X_j) = \boldsymbol{y}$ ($\boldsymbol{y}$ は親ノード群の各値で構成したベクトル) のもとでの n 通りの離散状態を持つ変数 X_j の条件付き確率分布を $p(X_j = x_1|\boldsymbol{y}), \cdots, p(X_j = x_n|\boldsymbol{y})$ とする (ただし $\sum_{i=1}^{n} p(x_i|\boldsymbol{y}) = 1.0$)．これを行として，親ノードがとりえる全ての可能な状態 $\pi(X_j) = \boldsymbol{y_1}, \cdots, \boldsymbol{y_m}$ について列を構成した表 1 が X_j にとっての CPT，$P(X_j|\pi(X_j))$ である．

表 1: 条件付き確率表 (CPT)

$p(X_j = x_1\|\pi(X_j) = \boldsymbol{y_1})$	$\cdots$	$p(X_j = x_n\|\pi(X_j) = \boldsymbol{y_1})$
:	$\ddots$	:
$p(X_j = x_1\|\pi(X_j) = \boldsymbol{y_m})$	$\cdots$	$p(X_j = x_n\|\pi(X_j) = \boldsymbol{y_m})$

また連続的な条件付き確率を表す場合には，確率密度関数をパラメータを持つ関数で定義するパラメトリックな確率分布で定義する方法がある．しかし連続変数のベイジアンネットはモデルとして表現することはできても，次で説明する確率推論を実行する良いアルゴリズムが離散の場合ほど確立されていないために，まだまだ課題が多く残されている．

4 ベイジアンネットの使用法

ベイジアンネットワークを使用するためには，適用する問題に適した (i) 重要な変数の選択, (ii) 適当なグラフ構造の設計, (iii) 条件付き確率の決定, の作業が必要であり，問題領域における十分なデータの収集と解析が重要である．また問題が複雑になりデータ量が増大するにつれ，以上の作業を全て人手で行なうことはそれほど容易ではなくなる．そこで様々な問題に応じたベイジアンネットワークを自動的に構成することを目指して，これらを統計データから学習により行う方法が研究されている．ここでは，変数はすでに与えられているものとして，条件付き確率の学習, グラフ構造の学習，モデルを構築した後の確率推論の順に説明する．

4.1 条件付き確率の学習

学習に用いるデータセットが確率変数の取り得る全ての組合せについての事例を含んでいる場合は完全データと呼ばれ，この場合には CPT の全ての項を埋めることができる．簡単のため確率変数が真偽二値とすると，親ノード群 $\pi(X_j)$ がある値をとる全ての事例数を N とし，さらにそのうち X_j が真であった事例数を n とする．仮に真の確率が $P(X_j = 1|\pi(X_j)) = \theta^*$ であったとすると，この観測結果が得られる確率は ${}_N C_n \theta^{*n}(1-\theta^*)^{N-n}$ になる． この θ^* を n, N から推定したい．データの数が多く，N が十分大きい場合には，漸近一致性を持つ最尤推定量 $\tilde{\theta} = n/N$ を点推定量として使う．ただし事例数が少ないときには最尤推定量と真の確率の値がそれほど近くならない場合がある．その場合には点推定量ではなくベイズ的に θ^* として想定できる確率分布を考える．具体的にはこれが n, N から次のベータ事前分布 (一般の離散多値の変数の場合は Dirichlet 事前分布)

$$\frac{1}{B(n+1, N-n+1)} \theta^n (1-\theta)^{N-n},$$

($B(n+1, N-n+1)$ はベータ関数) から定まると考える [36, 17, 15]．確率変数が連続値である場合は，先に述べたように条件付き確率分布がパラメトリックな関数として表わされるため，条件付き確率の学習はデータからのパラメータ推定の問題に帰着される．

データが全ての起こり得る組合せを含まなかったり，観測されない変数 (隠れ変数) を含む場合には不完全データと呼ばれる．このときにはまず，未観測データについての確率分布を推定し，さらにその分布によって期待値計算を行なうことが考えられる．この計算のためにマルコフ連鎖モンテカルロ法や EM アルゴリズム [11] などの手法が適用されている．

4.2 グラフ構造の学習

以上で述べた条件付き確率の学習はある特定のグラフ構造のもとでのパラメータ探索であった．一方，このグラフ構造もデータから決定したいという要求がある．データから構造を評価するときはその構造のもとでの最適なパラメータが必要であり，したがってグラフ構造の学習はパラメータの探索を含む．最適なパラメータへの収束が必ずしも保証できない山登り法などで探索する場合には準最適なパラメータで代用せざるを得ない．そのため構造の評価を正確に行うことが難しくなるという問題がある．またパラメータの探索空間はパラメータの次元のオーダーであるが，構造の探索空間はグラフのノード数に対して指数オーダーとはるかに広大なため，データセットだけから最適な構造を発見することは容易でない．こうした様々な困難のためグラフ構造の学習は現在も未解決の多くの問題を含んでおり，今後もなお重要な課題であると認識されている [13]．

現在，良く知られているベイジアンネットワークの構造学習アルゴリズムとしては現実的な時間でグラフ構造を探索するためのヒューリスティクスを用いた K-2 アルゴリズム [7] がある．このアルゴリズムは (1) 各ノードについて親ノードになりえる候補を限定しておく，(2) ある子ノードを一つ選び，候補となる親ノードを一つづつ加えてグラフを作る，(3) そのグラフのもとでパラメータを決定し，評価する，(4) 評価が高くなった時だけ親ノードとして採用し，(5) 親ノードとして加える候補がなくなるか，加えても評価が高くならなくなったら他の子ノードへ移る，(6) 全ての子ノードについて (1)-(5) を繰り返す，というものである．始めにノードを順序づけしておくことで親ノードになれるノードの組合せを限定し，計算量の増大と生成するグラフ構造の循環を避けているが，これにより生成できる構造の探索範囲も制約されてしまう．他には順序づけをしないヒューリスティクスを用いた探索アルゴリズム [3] や遺伝的アルゴリズムを使って探索するもの [25] などがある．

生成したグラフ構造を評価する際には，過剰に複雑な構造を避けるために，データの尤度と構造の複雑さの両方を考慮することが本質的に重要であり，モデルを選択するために MDL 基準 [37] やその他様々な情報量基準が使用されている．

4.3 確率推論

構築したベイジアンネットの上での確率計算，未知の変数についての予測を行う確率推論については本チュートリアルにて鈴木氏が詳説するので，ここでは簡単に触れておく．

ベイジアンネットワークによる確率推論は観測された変数の確定値 (evidence:e) から，知りたい確率変数 (X) のとりえる具現値の確率，すなわち事後確率分布 $P(X|e)$ を求め，それにより変数の期待値や確信度最大の具現値 (MAP 値)，複数の変数によって表されるある仮説の確信度 (いくつかの変数が特定の値の組をとる場合の同時確率) などを評価することである[3]．

すなわち，i) 観測された変数の値 e をノードにセットする， ii) 親ノードも観測値も持たないノードに事前確率分布を与える， iii) 知りたい対象の変数 X の事後確率 $P(X|e)$ を得る，という手順で計算を行なうことがベイジアンネットワークの上の確率推論である．

iii) における事後確率を求めるために，evidence からの確率伝播 (変数間の局所計算) によって各変数の確率分布を更新していく belief propagation と呼ばれる計算がベイジアンネットワークの上の計算アルゴリズムの特徴である．ここでは簡単な図 3 の構造での計算の例を示す．

$X_j \rightarrow X_{j+1}$ の条件付き依存性が成立しており，計算するノードから先の親ノードに与えられる evidence を e^+, 計算するノードから先の子ノードに与えられる evidence を e^- とする．ベイズの定理より，ノード X_j の確率は

$$\begin{aligned} P(X_j|e) &= P(X_j|e^+, e^-) \\ &= \frac{P(e^-|X_j, e^+)P(X_j|e^+)}{P(e^-|e+)}. \end{aligned}$$

$\alpha = \frac{1}{P(e^-|e+)}$ を X_j の値によらない正規化定数とし，e^+ と e^- は X_j が与えられると条件付き独立になるので，

$$P(X_j|e) = \alpha P(e^-|X_j)P(X_j|e^+). \qquad (3)$$

となる．このうち e^+ による X_j への寄与分,$P(X_j|e^+)$ は親ノードの確率，$P(X_{j-1}|e^+)$ と X_j の CPT を使った周辺化とよぶ計算, 式 (4) によって求めることができる．

$$P(X_j|e^+) = \sum_{X_j} P(X_j|X_{j-1})P(X_{j-1}|e^+). \qquad (4)$$

[3]これは推定対象をパターンクラス，観測変数をパターン信号とすると，Bayesian classifier と呼ばれるものであり，MAP 値を取るということはベイズ最適な識別を行うことに等しい．したがってベイジアンネットのグラフ構造を積極的に活用して認識対象をモデル化することで，そのモデルに基づくパターン認識が実現できる．

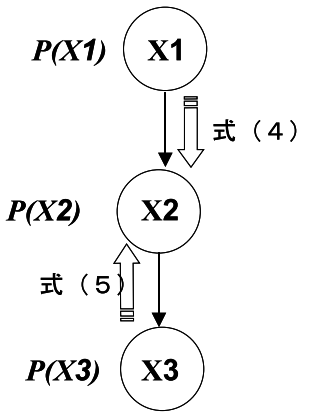

図 3: ベイジアンネットワーク上の確率伝播

$P(X_{j-1}|e^+)$ は式 (4) を再帰的に適用して次の親ノードと CPT から求まる．最終的には evidence ノードでは e^+，それより先に親ノードが存在しないノードの場合は事前確率分布を代入する．

一方，子ノード側の e^- の寄与分，$P(e^-|X_j)$ を計算するためには，X_{j+1} に関する条件付き確率 $P(X_{j+1}|X_j)$ を使った周辺化を行なう．

$$P(e^-|X_j) = \sum_{X_{j+1}} P(e^-|X_j, X_{j+1})P(X_{j+1}|X_j)$$

ここで，X_{j+1} を固定した時には，X_j より下にある e^- は X_j の値によらず独立であるから，

$$P(e^-|X_j) = \sum_{X_{j+1}} P(e^-|X_{j+1})P(X_{j+1}|X_j). \quad (5)$$

ここで，$P(e^-|X_{j+1})$ は，式 (5) を再帰的に適用していき，CPT と次の子ノードより求まる．最終的には evidence ノードでは観測値，それより下に子ノードを持たない最終のノードでは一様分布を代入する．したがって，以上式 (4),(5) を式 3 に代入して掛け合わせ α により正規化することで，各ノード X_j の事後確率 $P(X_j|e)$ が求まる (図 3)．

この局所的な変数間を伝播するような確率計算が繰り返し行なわれ，観測値として確定値の代入された evidence を起点にしてネットワーク全体をくまなく伝播することで各ノードの確率値が得られる．

ただし，一般的なグラフの場合にはノードには複数の親，子ノードが存在し，また観測値 e が多数入ることもある．そのため，計算の複雑性はグラフの構造や観測値の入り方により大きく影響を受ける．とくに計算はグラフ上のリンクをたどって行われるため，パス[4]のたどり方が計算効率の上で重要である．任意の二つのノード間にパスが一つしか存在しない singly connected グラフの場合には最悪計算量がノードの数にたいしては線形オーダーとなる効率の良い確率伝播アルゴリズムが複数の研究者により提案されている [30, 26]．

ノード間のパスが複数存在してしまう一般のグラフ構造, multiply connected グラフの場合には，最悪計算量が NP 困難になるという問題がある [6]．そこで，この場合には複数のノードを併合してグラフ構造を一つの singly connected グラフに変換する (clustering)，確率変数を複数の値に場合分けして，複数の singly connected グラフに変換する (conditioning)，モンテカルロサンプリングを行う，などの様々な近似的な解決法が考えられている．とくに最近では，Clustering アルゴリズムの一種として junction tree アルゴリズムが計算効率の良いアルゴリズムとして定評を得ている．これは共通の親ノードを併合する操作を繰り返してノードのクリークをクラスタとして生成し，元のベイジアンネットを クリークを結合した木構造からなる Moral graph と呼ばれる無向グラフに変換する．この木構造にしたがってクリーク毎に効率的な確率計算を行う．グラフ変換にかかる計算コストは比較的大きくなるが，一度変換した後に何度も確率計算を行う場合には非常に効率が良い．また近似計算による効率的なアルゴリズムとしては variational method(変分法) [5]と平均場近似の応用 [21, 40, 38] も研究されている．

5　ベイジアンネットの拡張

これまで離散変数のベイジアンネットワークを中心に解説してきたが，実際の問題に適用する場合にはさらに多様な変数を扱う必要があり，それにともなって，モデルとして多くのバリエーションがありえる．ベイジアンネットによるモデリングの一つの特長はこうしたネットワークと条件付き確率による表現が高い記述性と一般性を持つことである．したがって自然な拡張として変数や条件付きパラメータの表現方法の拡張や機能素子の追加によってより幅広い問題領域に対応することが可能となる．

例えばノイズを含むアナログセンサからの観測値などは連続確率変数になるので，離散確率変数のかわりに連続確率変数でベイジアンネットワークを構成する．この場合には条件付き確率は連続的な関数によって表す必要があり，適当なパラメータを導入して特定の関数族を仮定することで表現する．これまでの例ではとくにガウス分布が用いられることが多く，このように構成したベイジアンネットワークはガウシアンネットワーク [14] ある

[4]ここでいうパスとはリンクの向きを無視した接続関係のこと．

[5]例えば尤度関数の計算が容易でない場合にそれより計算の簡単な式で代替し，尤度の上限などを評価する近似計算．

いは Conditional Gaussian[27] と呼ばれている．例えば親ノード $X = x$, 子ノード $Y = y$ とすると，条件付き確率は，

$$P(y|x) = g(x) \cdot e^{-f(x,y)} \tag{6}$$

のように書ける (f, g は適当な関数として与える f は通常は X に関して線形の関数である．さらに変数間の関係が複雑な場合には非線形な関数を導入する必要があり，例えば階層型ニューラルネットワークを適用することも考えられる [29]．

また時間遅れ (状態遷移) を持つ系をモデル化する場合のベイジアンネットワークとして Dynamic Belief Network(DBN) と呼ばれるものがある [9]．時系列を扱う確率モデルとして工学的に重要なマルコフチェイン，あるいはその発展形であり，音声認識や遺伝子情報処理で使われている隠れマルコフモデル (Hidden Markov Model:HMM)[6][32] があるが，このような時間的な遷移をモデル化するために，DBN はベイジアンネットワークの一部の有向リンクに状態遷移と時間順序の意味を持たせ，時間軸に沿って必要な回数だけ繰り返し適用することで，時刻パラメータ t を持つ有限個の (独立かつ同一分布に従う) 確率変数 $\{X_1(t), \ldots, X_N(t)\}$ の分布を表現する．隠れマルコフモデルは状態遷移と出力として確率を割り当てるのに対し，DBN は全てのノードに対する条件付き確率として割り当てる点が異なり，ある時刻の状態は任意の複数のノードの組で表せる．

他にも無向グラフと有向グラフを組み合わせた連鎖グラフ (chain graph) や，グラフの中に意志決定を行なう計算ノードを導入した inference diagram, decision network[34] などもベイジアンネットワークのバリエーションとして考えることができる．こうした拡張の一つとして複雑なシステムの制御，意志決定を実現するために，エージェントアーキテクチャの中の一部としてベイジアンネットを利用する試みを行っている [45](図 4)．これは状況に依存するエージェント群でシステムを構成し，ベイジアンネットによる確率的な状況認識を行うことで最適な行動選択を行う．

またベイジアンネットワークの表現の直観性，計算の能率性が明らかになる一方で，本質的な表現能力の限界についての問題意識もある．論理構造としては命題論理レベルの表現であることと変数が有限次元であるという制約から，一階述語論理レベルの表現との統合にも目が

[6]隠れマルコフモデルは入力データの長さに比例した実行時間で済む Baum-Welch EM 学習アルゴリズムや Viterbi 最尤遷移推定アルゴリズムなど優れた確率計算アルゴリズムを備えており，音声認識，遺伝子情報処理など幅広い範囲でその有用性が実証されている確率モデルである．

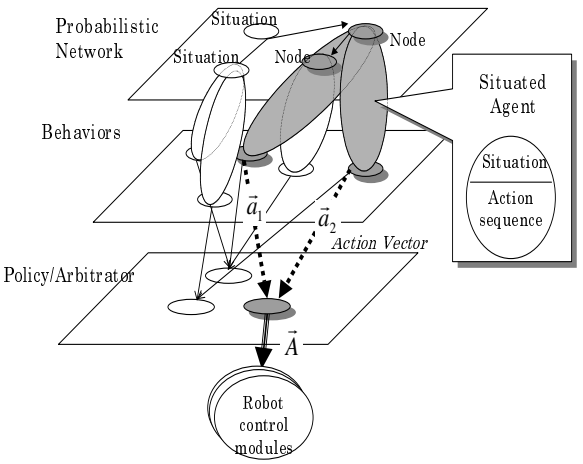

図 4: 確率ネットによるマルチエージェントアーキテクチャ

向けられている (これについては本チュートリアルでは佐藤氏による解説が行われる)．とくに問題領域について事前知識がわかっている場合にはこれを論理的に表現してモデル構築に利用する方法 [2, 31, 24, 35] も有望と考えられる．

6 ベイジアンネットの応用

確率理論に基づく厳密な確率計算が容易でなかった時代には初期のエキスパートシステム MYCIN で使用された確信度表現，ファジィ理論などの比較的簡単な計算で不確実性を扱える手法などが提案されてきた．しかし十分な計算能力が得られるようになった今では，数学的に厳密な確率理論に基づき，より意味づけの正確な尤度や期待値などの統計量を得ることも重要である．ベイジアンネットワークの出力は該当する変数についての確率分布, すなわち変数がとる全ての可能な状態について確率値を割り当てた分布であり，数理的意味が明らかな，変数の期待値，事後確率最大値，エントロピー，相互情報量などを計算できることが一つのメリットでもある．それにより期待効用の最大化や，情報量最大化を実現する操作を行うことができ，不確実な状況におけるシステムの合理性[7]を保証することが可能になる．

国際的には実際の応用例も増え [16]，今年で 17 回になる国際会議 Uncertainty in AI を中心とし，人工知能，統計，機械学習，ニューラルネット，情報理論など分野を越えて国際的な広がりを見せている[8]．90 年代後半に

[7]期待効用を最適化するという意味での合理性である [34]．

[8]国際的にはどの分野でもベイジアンネットの認知度は日本よりもはるかに高い

なって，高速な計算法である Junction Tree Algorithm, その実装である Hugin[19] が現れたことで，性能評価も容易になり，企業での本格的な応用も進んでいる．以降，ベイジアンネットを実際に適用した国内外の具体的な事例についていくつか触れておく．

以前からベイジアンネットワークの基礎研究を進めていたマイクロソフトは，同社の商用ソフトウェアである Office に確率推論を行なうユーザモデル (Bayesian User Model)[18] を応用していると述べている．アプリケーションソフトを使用中のユーザの挙動の観測から，2 の例のようなベイジアンネットワークを構築し，ユーザのマウス操作の履歴 (どのメニューをたどってきたか，反応時間が早いか遅いかなど)，ソフトの実行可能な状態などからユーザが次に行おうとしているコマンドやゴール状態を推定し，現在のユーザが必要としていると思われる適切な助言を生成する[9]．さらにその最新バージョンである Office XP ではさらに積極的に確率推論機能を利用している．

秋葉らは過去のユーザの発話からユーザが対象についてどの位知っているかを推定するベイジアンネットワークを用いたユーザモデルを内蔵して渋谷の道案内を行なう連続音声認識対話システムを実現した [1]．こうしたユーザモデリングへのベイジアンネットの適用 [33, 41] はデータの整備と統計的学習の発達によって今後ますます有望になると思われる．また自然言語による対話のプランニングにもベイジアンネットが適用されており [5], 特定のタスクにおける対話プランニングへの実装例もある [43]．今後対話コーパスが普及するにつれて対話制御へのさらなる応用も期待される (これに関連した話題は本チュートリアルで秋葉氏により説明される)．

対話のプランニングと類似した問題構造を持つロボットや自律移動システムのコントロール (これについては本チュートリアルでは稲邑氏が実例をあげて説明する) も典型的な応用例である [10]．カルフォルニア大バークレイの S.Russell らのグループでは，高速道路での自動車運転シミュレータを用いて，運転中の直前の状態から次の状態を予測する Dynamic Belief Network を構築し，自動運転システムのフィージビリティスタディを行った [12]．

他にも大量のデータが得やすい問題への適用は比較的有効であり豊富な医療データに基づいた，筋神経系，リンパ節，先天性心臓疾患 [20]，貧血症 [23] などの医療診断システムも知られている (遺伝子情報処理に関連する話題は本チュートリアルにおいて有田氏から紹介がある)．

また通信理論分野で最近脚光を浴びている Turbo 符号化の周辺でもベイジアンネットワークの確率伝播アルゴリズムと復号化アルゴリズムとの関連が議論されている [28, 22](これについては本チュートリアルで松嶋氏による解説がある)．

7　まとめ

情報技術にたいする社会的，経済的要請にともなって，環境や人間に内在する様々な不確定性を取り扱うモデリング技術は以前とくらべはるかに重要となっている．そのためにベイジアンネットを使った情報表現によるモデル化と，その応用の可能性に期待をしている．もちろん人間の認知過程，経済現象や生命現象のような複雑な確率事象のモデル化と予測は如何なる手法を用いてもけっして容易な問題ではなく，またベイジアンネットワークも可能な数多くのアプローチの内の一つに過ぎない．しかしながら，ベイジアンネットワークが他の多くの確率モデルのより一般的な表現となっており，包括的な視点を与えることができることや，また確率推論や学習を行うソフトウェアとして実用化も進んで来ていることから，これを不確実な対象に対する確率的知識表現の一つの道具として積極的に利用していくことは多くの未開拓の問題領域にアプローチし，研究課題を統一的に議論していく上でも効率的な手段となりえるだろう．海外でも最近とくにその理解と利用が進んできている．今回わが国でも本格的なチュートリアルを開催し，幅広い研究領域に属する研究者が一同に会した意義は大きいと言えよう．今後，国内で関連研究の意見交流，議論を行う場として，こうした機会を整える努力をしていくことも重要であろう．

[9] ただし，製品化にあたって，簡易的な実装にとどめたこと，適応性が低下したことから本来の意図通りの成果は上げていなかったようである．

参考文献

[1] Akiba, T. and Tanaka, H.: A Bayesian approach for user modelling in dialogue systems, *Proc. of the International Conference on Computational Linguistics*, pp. 1212–1218 (1994).

[2] Breese, J.: Construction of Belief and Decision Networks, *J. of Computational Intelligence*, Vol. 8, No. 4, pp. 624–647 (1992).

[3] Buntine, W.: Theory refinement on Bayesian networks, *Proc. of the 7th Conference on Uncertainty in Artificial Intelligence*, pp. 52–60 (1991).

[4] Castillo, E., Gutierrez, J., and Hadi, A.: *Expert Systems and Probabilistic Network Models*, Springer-Verlag (1997).

[5] Charniak, E. and Goldman, R.: A Bayesian model of plan recognition, *Artificial Intelligece*, Vol. 64, pp. 53–79 (1993).

[6] Cooper, G. F.: The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks, *Artificial Intelligence*, Vol. 42, pp. 393–405 (1990).

[7] Cooper, G. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from Data, *Machine Learning*, Vol. 9, pp. 309–347 (1992).

[8] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J.: *Probabilistic Networks and Expert Systems*, Springer-Verlag (1999).

[9] Dean, T. and Kanazawa, K.: A model for Reasoning about Persistence and Causation, *Computational Intelligence*, Vol. 5, No. 3, pp. 142–150 (1989).

[10] Dean, T. and Wellman, P.: *Planning and Control*, Morgan Kaufmann (1991).

[11] Dempster, A., Laird, N., and Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of Royal Statististical Society B*, Vol. 39, pp. 1–38 (1977).

[12] Forbes, J., Huang, T., Kanazawa, K., and Russel, S.: The BATmobile: Towards a Bayesian Automated Taxi, *Proc. of the 14th International Joint Conference on Artificial Intelligence*, pp. 1878–1885 (1993).

[13] Friedman, N., Goldszmidt, M., Heckerman, D., and Russell, S.: Challenge: Where is the Impact of Bayesian Networks in Learning?, *Proc. of the 15th International Joint Conference on Artificial Intelligence*, pp. 10–15 (1997).

[14] Geiger, D. and Heckerman, D.: Learning Gaussian Networks, *Proc. of the 10th Conference on Uncertainty in Artificial Intelligence*, pp. 235–243 (1994).

[15] Geiger, D. and Heckerman, D.: A characterization of the Dirichlet distribution with application to learning Bayesian networks, *Proc. of the 11th Conference on Uncertainty in Artificial Intelligence*, pp. 196–207 (1995).

[16] Haddawy, P.: An Overview of Some Recent Developments in Bayesian Problem Solving Techniques, *AI Magazine special issue on Bayesian Techniques Plus...*, Vol. 20, No. 2, pp. 11–20 (1999).

[17] Heckerman, D., Geiger, D., and Chickering, D.: Learning Bayesian networks: the combination of knowledge and statistical data, *Machine Learning*, Vol. 20, pp. 197–243 (1995).

[18] Horvitz, E., Breese, J., Heckerman, D., Hovel, D., and Rommelse, K.: The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users, *Proc. of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 256–265 (1998).

[19] Hugin expert, http://www.hugin.com (2001).

[20] Jensen, F.: *An Introduction to Bayesian Networks*, University College London Press (1996).

[21] Jordan, M., Ghahramani, Z., Jaakkola, T., and Saul, L.: An Introduction to Variational Methods for Graphical Models, *Learning in Graphical Models*, pp. 105–161 (1998), Kruwer Academic Publisher.

[22] Kabashima, Y. and Saad, D.: Belief Propagetion vs. TAP for decoding corrupted messages, *Europhys. Letter*, Vol. 44, No. 5, pp. 668–674 (1999).

[23] Kappen, H. J. and al., et : Approximate inference in medical diagnosis, *Pattern Recognition Letters* (1999).

[24] Koller, D. and Pfeffer, A.: Learning probabilities for noisy first-order rules, *Proc.of IJCAI'97, Nagoya*, pp. 1316–1321 (1997).

[25] Larrañaga, P., Poza, M., Yurramendi, Y., Murfa, R., and Kujipers, C.: Structure learning of Bayesian networks by genetic algorithms: a performance analysis of control parameters, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, pp. 912–926 (1996).

[26] Lauritzen, S. and Spiegelhalter, D.: Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society B*, Vol. 50, pp. 157–224 (1988).

[27] Lauritzen, S. and Wermuth, N.: Graphical Models for Associations Between Variables, Some of which are Qualitative and Some Quantitative, *Anals. of Statistics*, Vol. 17, pp. 31–57 (1989).

[28] Mackay, D.: Turbo Decoding as an Instance of Pearl's Belief Propagation Algorithm, *submitted to IEEE Journal on Selected Areas in Communication* (1996).

[29] Motomura, Y. and Hara, I.: Bayesian Network Learning System based on Neural Networks, *the proc. of int. symp. on Theory and Application of Softcomputing 2000* (2000).

[30] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, CA (1988).

[31] Poole, D.: Probabilistic Horn abduction and Bayesian networks, *Artificial Intelligence*, Vol. 64, pp. 81–129 (1993).

[32] Rabiner, L. and Juang, B.: *Foundations of Speech Recognition*, Prentice-Hall (1993).

[33] Rosis, F., et.al.: Modelling the User Knowledge by Belief Networks, *Journal of User Models and User-Adapted Interaction*, vol.2, No.4, pp.367–388, (1992).

[34] Russell, S. and Norvig, P.: *Artificial Intelligence, A modern approach*, Prentice Hall (1995), 邦訳: 古川康一監訳，エージェントアプローチ人工知能，共立出版 (1997).

[35] Sato, T. and Kameya, Y.: A Viterbi-like algorithm and EM learning for statistical abductuion, *to be presented at the UAI-2000 workshop on Fusion of Domain Knowledge with Data for Decision Support* (2000).

[36] Spiegelhalter, D. J., Lauritzen, S. L., Dawid, A. P., and Cowell, R. G.: Bayesian analysis in expert systems, *Statistical Science*, Vol. 8, pp. 219–247 (1993).

[37] Suzuki, J.: A construction of Bayesian networks from databases based on an MDL principle, in *Proc. of the 9th Conference on Uncertainty in Artificial Intelligence*, pp. 266–273 (1993).

[38] Tanaka, T.: A theory of Mean Field Approximation, *Advances in Neural Information Processing Systems*, Vol. 11, pp. 351–357 (2000), MIT Press.

[39] Whittaker, J.: *Graphical Models in Applied Multivariate Statistics*, John Wiley and Sons (1990).

[40] Wiegerinck, W. and Barber, D.: Mean Field Theory based on Belief Networks for Approximate Inference, *Proc. of the International Conference on Artificial Neural Networks*, pp. 499–504 (1998).

[41] 植野, 大西, 繁桝: 確率ネットワークを組み込んだテスト理論の提案, 電子情報通信学会論文誌, 77-A, 10, pp.1398–1408, (1994).

[42] 石塚満 (訳) : 15 章: 確率的推論システム, S.Russell and P.Norvig 著, 古川康一監訳，エージェントアプローチ人工知能, pp. 439–473 (1997), 共立出版.

[43] 乾, 徳永, 田中 : 意志決定理論に基づく 発話プランニング, 人工知能学会論文誌, Vol. 12, No. 5, pp. 760–769 (1997).

[44] 宮川雅巳:グラフィカルモデリング, 朝倉書店 (1998).

[45] 本村, 原: 確率ネットワークに基づくマルチエージェントアーキテクチャ, 第 28 回 情報処理学会数理モデル化と問題解決研究会 (2000).